



## King's Research Portal

DOI:

[10.1007/s00787-014-0593-0](https://doi.org/10.1007/s00787-014-0593-0)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Pina Camacho, L., Garcia-Prieto, J., Parellada, M., Castro-Fornieles, J., Gonzalez-Pinto, A. M., Bombin, I., Graell, M., Paya, B., Rapado-Castro, M., Janssen, J., Baeza, I., Pozo, F. D., Desco, M., & Arango, C. (2015). Predictors of schizophrenia spectrum disorders in early-onset first episodes of psychosis: a support vector machine model. *European child & adolescent psychiatry*, 24(4), 427-440. <https://doi.org/10.1007/s00787-014-0593-0>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Predictors of schizophrenia spectrum disorders in early-onset first episodes of psychosis: a support vector machine model

Laura Pina-Camacho<sup>1,2,\*</sup>

Email [lpina.iisgm@gmail.com](mailto:lpina.iisgm@gmail.com)

Juan Garcia-Prieto<sup>3</sup>

Mara Parellada<sup>1</sup>

Josefina Castro-Fornieles<sup>4,5</sup>

Ana M. Gonzalez-Pinto<sup>6</sup>

Igor Bombin<sup>7,8</sup>

Montserrat Graell<sup>9</sup>

Beatriz Paya<sup>10</sup>

Marta Rapado-Castro<sup>1,11</sup>

Joost Janssen<sup>1</sup>

Inmaculada Baeza<sup>4,5</sup>

Francisco Del Pozo<sup>3</sup>

Manuel Desco<sup>12,13</sup>

Celso Arango<sup>1</sup>

<sup>1</sup> Child and Adolescent Psychiatry Department, Hospital General Universitario Gregorio Marañón. School of Medicine, Universidad Complutense, IiSGM, CIBERSAM, Ibiza 43, 28009 Madrid, Spain

<sup>2</sup> Department of Child and Adolescent Psychiatry, Institute of

Psychiatry, King's College London, De Crespigny Park, SE5  
8AF London, UK

<sup>3</sup> Laboratory of Cognitive and Computational Neuroscience, Centre for Biomedical Technology (CTB), Universidad Politécnica de Madrid, Campus De Montegancedo, Ctra. M-40, Km. 38, Pozuelo De Alarcón, 28223 Madrid, Spain

<sup>4</sup> Child and Adolescent Psychiatry and Psychology Department, Neuroscience Institute, Hospital Clínic, IDIBAPS, SGR-1119 Barcelona, Spain

<sup>5</sup> Department of Psychiatry and Psychobiology, University of Barcelona, CIBERSAM, Villarroel 170, 08036 Barcelona, Spain

<sup>6</sup> International Mood Disorders Research Centre. Hospital Santiago Apóstol, University of the Basque Country, CIBERSAM, Olaguibel 29, 01004 Vitoria, Spain

<sup>7</sup> Reintegra: Neuro-Rehabilitation Center, Eduardo de Fraga Torrejón 4, 33011 Oviedo, Spain

<sup>8</sup> Department of Psychology, University of Oviedo, CIBERSAM, 33003 Plaza Feijóo, Oviedo, Spain

<sup>9</sup> Section of Child and Adolescent Psychiatry and Psychology, Hospital Infantil Universitario Niño Jesús, CIBERSAM, Av de Menéndez Pelayo 65, 28009 Madrid, Spain

<sup>10</sup> Child and Adolescent Mental Health Unit, Department of Psychiatry and Psychology, Hospital Universitario Marqués de Valdecilla, CIBERSAM, Av Valdecilla, s/n, 39008 Santander, Spain

<sup>11</sup> Melbourne Neuropsychiatry Centre, The University of Melbourne, Melbourne, Australia

<sup>12</sup> Department of Experimental Medicine, Hospital General Universitario Gregorio Marañón. IiSGM, CIBERSAM, Ibiza 43, 28009 Madrid, Spain

<sup>13</sup> Bioengineering and Aerospace Engineering Department, Universidad Carlos III de Madrid, Av de la Universidad

## Abstract

Identifying early-onset schizophrenia spectrum disorders (SSD) at a very early stage remains challenging. To assess the diagnostic predictive value of multiple types of data at the emergence of early-onset first-episode psychosis (FEP), various support vector machine (SVM) classifiers were developed. The data were from a 2-year, prospective, longitudinal study of 81 patients (age 9–17 years) with early-onset FEP and a stable diagnosis during follow-up and 42 age- and sex-matched healthy controls (HC). The input was different combinations of baseline clinical, neuropsychological, magnetic resonance imaging brain volumetric and biochemical data, and the output was the diagnosis at follow-up (SSD vs. non-SSD, SSD vs. HC, and non-SSD vs. HC). Enhanced recursive feature elimination was performed for the SSD vs. non-SSD classifier to select and rank the input variables with the highest predictive value for a diagnostic outcome of SSD. After validation with a test set and considering all baseline variables together, the SSD vs. non-SSD, SSD vs. HC and non-SSD vs. HC classifiers achieved an accuracy of 0.81, 0.99 and 0.99, respectively. Regarding the SSD vs. non-SSD classifier, a combination of baseline clinical variables (severity of negative, disorganized symptoms and hallucinations or poor insight) and neuropsychological variables (impaired attention, motor coordination, and global cognition) showed the highest predictive value for a diagnostic outcome of SSD. Neuroimaging and biochemical variables at baseline did not add to the predictive value. Thus, comprehensive clinical/cognitive assessment remains the most reliable approach for differential diagnosis during early-onset FEP. SVMs may constitute promising multivariate tools in the search for predictors of diagnostic outcome in FEP.

AQ1

---

## Keywords

Early-onset psychosis

Schizophrenia and disorders with psychotic features

Diagnosis

Child and adolescent psychiatry

## Electronic supplementary material

The online version of this article (doi:10.1007/s00787-014-0593-0) contains supplementary material, which is available to authorized users.

## Introduction

Psychotic disorders are among the leading causes of disease burden in adolescents and young people [1, 2]. The rapid distinction between schizophrenia spectrum disorders (SSD) and other psychotic disorders such as bipolar disorder is thus an important challenge for child and adolescent psychiatrists, since it may enable early optimization of treatment and accurate use of prognostic statements [3]. However, this could prove extremely difficult in patients with first episodes of psychosis (FEP), particularly in early-onset cases, for a number of reasons. First, although the clinical approach to psychosis is evolving towards a dimensional model [4], symptom-based categorical classification of patients is still the basis for diagnosis during a FEP [5, 6]; however, differential diagnosis might be hampered by the lack of specificity of symptoms at this stage (e.g. positive psychotic symptoms may be shared by SSD, bipolar disorder and other psychotic disorders). Second, categorical classifications require a temporal criterion for a diagnosis of SSD or other psychotic disorders [5, 6]. Therefore, the initial diagnosis can shift during follow-up [7–9]. For these reasons, the stability of diagnosis in FEP patients is very low (around 60 %) [7–10]. Third, SSD and other disorders such as bipolar disorder with psychotic symptoms share not only symptomatic features but also many cognitive and biological features and risk factors [11].

To date, numerous clinical and biological variables have been studied during the FEP with regard to their diagnostic predictive value for SSD. For example, clinical variables such as history of obstetric complications [12], the presence of negative symptoms [9, 13], or poor general functioning [8, 14] at onset of the illness are good clinical predictors of an SSD diagnostic outcome both in adult and adolescent samples. Among neuropsychological variables, executive function impairments in FEP patients seem to be good predictors of schizophrenia both in adult [15] and adolescent samples [16]. Regarding neuroimaging variables, prefrontal, cingulate, insular and cerebellar gray matter (GM) volume reductions have been associated with

conversion to psychosis in individuals at high risk of psychosis [17], and GM deficits in the prefrontal cortex, insula, amygdala or hippocampus have been associated with a diagnosis of schizophrenia (but not bipolar disorder) both in FEP and chronic patients [18–20]. Lastly, low antioxidant status during early-onset FEP has been described both in patients with schizophrenia and bipolar disorder, while reduced plasma levels of glutathione were only found in those patients with schizophrenia [21].

Despite the increasing body of literature, the reported predictive weight of the above-mentioned variables has been too small to have diagnostic value for SSD [22, 23]. The main caveats of predictive studies to date include (1) the use of relatively small samples of patients that do not allow for subdivision of patients into diagnostic groups within psychosis, (2) the inclusion of chronic (instead of first-episode) patients, (3) the use of traditional multivariate statistical approaches such as linear discriminant analysis (LDA) or logistic regression (LR), and (4) the fact that most studies have focused on a specific measurement or a specific source of measurements (e.g. clinical, neuropsychological, neuroimaging or biochemical data alone), and have relied on a limited number of variables (e.g. total scores or subscores of a particular clinical or neuropsychological scale). Other fields of medicine have demonstrated that clinical prediction, especially in complex disorders, often requires a combination of multiple measurements from different sources [24]. To this end, prediction methodologies based on high-dimensional multivariate statistical approaches and including a combination of multiple clinical and biological data seem to be needed and could potentially improve prediction accuracy [25]. One of these methods is support vector machines (SVMs), a multivariate pattern recognition approach that emerges from the field of machine learning [26, 27], which enables the following: (1) classification of two diagnostic groups by considering high-dimensional qualitative and quantitative variable information from each subject, from different sources and at the same time, and (2) description of the combination of variables with the highest predictive value for each diagnostic group (i.e. those that contribute the most to the discriminating pattern from each diagnostic class). Compared to other multivariate pattern analysis techniques such as LR or LDA, SVMs require fewer variables to achieve better prediction estimates, perform better when high-correlation structures are observed in the data, do not need multiple comparison correction (e.g. false discovery rate [FDR] or Bonferroni correction), and do not impose any a priori assumptions on single variable

relevance or data distribution except that they be “independent and identically distributed” (iid) [28].

The aim of this study was to develop an SVM model to assess the differential predictive values for a diagnostic outcome of SSD in a large set of clinical, neuropsychological, neuroimaging and biochemical data at the emergence of an early-onset FEP. Based on previous literature, we proposed the following hypotheses: (1) a combination of clinical variables (negative symptoms, premorbid functioning), neuropsychological variables (executive function), and neuroimaging variables (frontal volume) at baseline (i.e. during the FEP) will help us predict a diagnostic outcome of SSD, and (2) the SVM model that includes this combination of variables at baseline will classify with high accuracy FEP patients who will eventually be diagnosed with SSD.

## Subjects and methods

### Procedure

The child and adolescent first-episode psychosis study (CAFEPS) is a 2-year, multicenter, prospective, longitudinal study of 110 patients with an early-onset FEP and 98 age- and sex-matched healthy controls who were recruited from six different sites in Spain. The inclusion criteria for patients were age between 7 and 17 years at the initial evaluation and a first psychotic episode following DSM-IV criteria [5] (the presence of positive symptoms such as delusions and/or hallucinations) of less than 6 months' duration. The exclusion criteria were the presence of a concomitant Axis I disorder, mental retardation according to DSM-IV criteria [5] (i.e. not only an intelligence quotient below 70, but also impaired functioning prior to onset of the disorder), pervasive developmental disorders, neurological diseases, history of head trauma with loss of consciousness, and pregnancy. The inclusion criteria for controls were similar age as patients, residence in the same geographical areas, and the absence of psychiatric or neurological disorders, head trauma, mental retardation, and pregnancy. Sample recruitment and methodology have been described elsewhere [29].

Diagnosis was made according to DSM-IV criteria [5] using the Spanish version of the Kiddie-SADS present and lifetime version (K-SADS-PL) instrument [30] at baseline and at years 1 and 2 by experienced child psychiatrists with specific training for the interview. For the purposes of this

study, patients were categorized as ‘SSD’ (diagnosed with schizophrenia, schizoaffective disorder, or schizophreniform disorder during follow-up) or ‘non-SSD’ (diagnosed with bipolar disorder, major depressive disorder with psychotic symptoms, brief psychotic disorder, or psychotic disorder not otherwise specified during follow-up). Other authors [9, 14] have grouped schizophrenia, schizoaffective disorder, and schizophreniform disorder together, as they all have similar schizophrenia-like psychotic features.

Clinical and neuropsychological assessments, a magnetic resonance imaging (MRI) scan and blood samples for oxidative stress determinations were obtained at baseline at all the sites. A complete description of the clinical, neuropsychological, neuroimaging and biochemical assessment procedures is provided as Online Resource 1.

## Selected sample and dataset construction

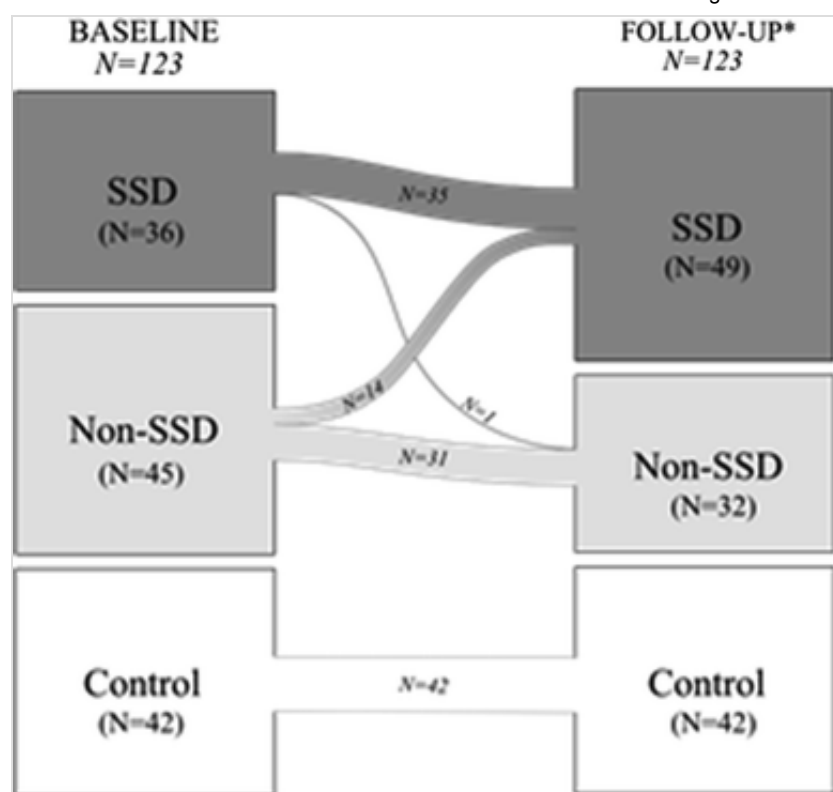
### Subject selection

Since the outcome variable of this study was a diagnostic outcome, only those CAFEPS patients with a stable K-SADS-based diagnosis at follow-up [i.e. obtained at year 1 ( $N = 18$ ) or year 2 ( $N = 63$ )] were included in the analysis. This was done to avoid the problem of diagnostic instability. In fact, 15 patients (18.5 % of patients included in this study) had a diagnostic category at baseline that shifted during follow-up (Fig. 1).

### Fig. 1

Diagnostic stability of first episodes of early-onset psychosis over the follow-up period. For all subjects, a diagnosis was established at baseline and at follow-up according to DSM-IV criteria [5] using the Kiddie-schedule for affective disorders and schizophrenia, present and lifetime version (K-SADS-PL) diagnostic scale [30]. Fifteen patients (14 non-SSD and 1 SSD) had a diagnostic category at baseline that shifted during follow-up. \*Follow-up refers to year 1 follow-up diagnostic assessment (available for 18 patients) or year 2 follow-up diagnostic assessment (available for 63 patients and 42 controls). *SSD* schizophrenia spectrum disorders (including schizophrenia, schizophreniform disorder, and schizoaffective disorder), *non-SSD* non-schizophrenia spectrum disorders (including bipolar disorder, major depressive disorder with psychotic symptoms, brief psychotic disorder, and psychotic disorder not otherwise specified)





Additionally, to include all the available information in the dataset on a subject-by-subject basis, only those subjects with complete or almost complete (>95 %) data for the four sets of variables (clinical, neuropsychological, neuroimaging and biochemical variables) were included in the analysis.

Out of the original CAFEPS sample of 110 patients, 10 were excluded because they did not have a stable K-SADS-based diagnosis at follow-up (since they were lost to follow-up before the visit at year 1), 18 were excluded because they did not have baseline imaging data, and 1 was excluded because of unavailable neuropsychological assessment at baseline. The final sample comprised 81 FEP patients (70 % males) with a mean age of  $15.38 \pm 1.90$  years (range 9–17 years). Of these, 49 patients were categorized as SSD [diagnosed with schizophrenia ( $n = 39$ ), schizoaffective disorder ( $n = 7$ ), or schizophreniform disorder ( $n = 3$ ) during follow-up], and 32 as non-SSD [diagnosed with bipolar disorder ( $n = 20$ ), major depressive disorder with psychotic symptoms ( $n = 5$ ), brief psychotic disorder ( $n = 2$ ), or psychotic disorder not otherwise specified ( $n = 5$ ) during follow-up]. To include size-balanced groups in the SVM models, we randomly selected 42 individuals from the original CAFEPS sample of 98 healthy controls. These subjects were age- and sex-matched with the SSD and non-SSD groups [mean age  $15.21 \pm 1.46$  years (range 12–17 years), 69 % males], with complete data available.

For the development of the SVM classifiers, the sample from the original dataset was divided into 2 sets: the training set [a randomly selected 80 % of the sample (39 SSD, 26 non-SSD and 34 HC)] and the test set [the remaining 20 % of the sample, with the same proportion of subjects belonging to the SSD ( $N = 10$ ), non-SSD ( $N = 6$ ), and HC ( $N = 8$ ) groups as in the training set].

Variable preprocessing

Before dividing the original dataset into two sets, preprocessing required all baseline variables defining each subject to be normalized and scaled to achieve an equalized histogram. Qualitative variables were scaled as binary matrices (e.g. for the sex variable, female [1,0], male [0,1]; for the family history of psychosis (first degree relative) variable: presence [1,0], absence [0,1]). Quantitative variables were scaled between 0 and 1 as follows: (1) when the variable had a defined range (e.g. PANSS scale 30–210), it was scaled as  $f(x) = (x - \min(x_i))/[\max(x_i) - \min(x_i)]$ , and (2) when there was no defined range (e.g. weight), it was normalized as

$$f(x) = x - [\text{mean}(x_i) - 5 \times (\text{standdev}(x_i))]/[\text{mean}(x_i) + 5 \times (\text{standdev}(x_i))] - [\text{mean}(x_i) - 5 \times (\text{standdev}(x_i))]$$

Variable selection and treatment of missing data

Table 1 summarizes the 1,050 baseline variables (747 clinical, 81 neuropsychological, 221 neuroimaging and 1 biochemical variable) included in the dataset.

1Q2

Table 1

Baseline assessment variables included in the dataset

Clinical variables ( $n = 747$ )
Demographic data
Age, sex, ethnicity, socioeconomic status, parent and child years of education
Developmental, medical and psychiatric records
Developmental history, educational history, personal medical and psychiatric history, family psychiatric history, substance misuse, psychopharmacological treatment, first psychotic symptoms, duration of illness
Clinical and functional scales <sup>a</sup>

Lewis–Murray scale of obstetric data, FES, PAS-childhood subscale, PANSS, HDRS, YMRS, SUMD—abbreviated version, CGI-S, CGAS, SCOS, SDQ—parent and child version, WHO-DAS—short version

Anthropometric data

Weight, height, BMI

Neuropsychological variables ( $n = 81$ )<sup>b</sup>

Neuropsychological battery

WISC-R/WAIS-III, TMT-A, TMT-B, Stroop test, TAVEC, FAS, COWAT, WCST, NES

Estimated IQ

—

Cognitive domains

Attention, speed of processing, learning and memory, working memory, executive function, global cognition score

Neuroimaging variables ( $n = 221$ )

ROIs<sup>c</sup>

Frontal lobe (whole frontal lobe, orbitofrontal cortex, dorsolateral prefrontal cortex), parietal lobe, temporal lobe (whole temporal, mesial and external subregions), occipital lobe, subcortical regions (caudate, putamen, thalamus, internal capsule, globus pallidus), hippocampus

Biochemical variables ( $n = 1$ )

Whole-brain GM, WM and CSF volumes, ICV

TAS

*BMI* body mass index, *CGAS* children's global assessment scale, *CGI-S* clinical global impression scale—severity, *COWAT* control oral word association test, *CSF* cerebrospinal fluid, *FAS* verbal fluency test, *FES* family environment scale, *GM* gray matter, *HDRS* Hamilton depression rating scale, *ICV* intracranial volume, *IQ* intelligence quotient, *NES* neurological evaluation scale, *PANSS* positive and negative symptom scale, *PAS* premorbid adjustment scale, *SCOS* Strauss–Carpenter outcomes scale, *SDQ* strengths and difficulties questionnaire, *SUMD* scale to assess unawareness of mental disorder, abbreviated version; *TAS* total antioxidant status, *TAVEC* Spanish version of the California verbal learning test; *TMT-A* and *TMT-B* trail making test, parts A and B, *WAIS-III* Wechsler adult intelligence scale, 3rd edition, *WCST* Wisconsin card sorting test, *WHO-DAS* World Health Organization disability assessment schedule, short version, *WISC-R* Wechsler intelligence scale for children-revised, *WM* white matter, *YMRS* young mania rating scale

<sup>a</sup>Ratings of every item, subscores (where appropriate) and total or summary scores (where appropriate) were included. A complete description of the clinical assessment procedure is provided as Online Resource 1

<sup>b</sup>For each test, raw scores and *z* scores of single test variables, raw and *z* subscores (where appropriate) and raw and *z* total or summary scores (where appropriate) were included. A complete description of the neuropsychological assessment procedure is provided as Online Resource 1

<sup>c</sup>For each ROI, raw volumes and volume percentages for each hemisphere and for GM, WM and CSF (within each ROI) were obtained. A complete description of the neuroimaging assessment procedure is provided as Online Resource 1

Variables for which  $\geq 15$  % of the subjects had no information available were excluded from the dataset. The following variables were therefore excluded: birth weight and Apgar score, puberal status, the Strauss–Carpenter outcomes scale (SCOS) ( $\geq 15$  % of the subjects had no information available for item number 6), the teacher version of the strengths and difficulties questionnaire (SDQ), the parent–adolescent communication inventory (PACI), the early adolescence, late adolescence, and adulthood subscales of the Premorbid adjustment scale (PAS), duration of the disability in the World Health Organization disability assessment schedule (WHO-DAS), raw and *z* scores of the continuous performance test-II (CPT), cellular glutathione peroxidase, catalase and superoxide dismutase activities, lipid peroxidation and levels of glutathione in plasma.

Variables for which  $< 15$  % of the subjects had no information available were included in the dataset. For these variables, missing values were treated as follows. Zero imputation was performed for missing qualitative values (missing values were expressed as empty matrices, [0, 0]). Subgroup mean imputation was performed for missing quantitative values (by replacing the missing value with the patient or control mean value for that variable, depending on subject's membership).

When a clinical variable was applicable only to patients [e.g. positive and negative symptom scale (PANSS) or Hamilton depression rating scale (HDRS)], values indicating the absence of symptoms or absence of impairment were imputed to healthy controls (e.g. rating of 1 for PANSS items, rating of 0 for HDRS items).

## Statistics

### Group differences at baseline

Group differences were tested for baseline demographic data (age, sex and

ethnicity), duration of illness, antipsychotic treatment, total PANSS score and children's global assessment scale (CGAS) score. Binary comparisons were performed for the following groups: (1) SSD vs. non-SSD, SSD vs. controls and non-SSD vs. controls included in the original dataset, (2) SSD vs. non-SSD, SSD vs. controls and non-SSD vs. controls included in the training set, (3) SSD vs. non-SSD, SSD vs. controls and non-SSD vs. controls included in the test set, (4) patients included in the training set vs. patients included in the test set, (5) controls included in the training set vs. controls included in the test set, (6) patients included and not included in the dataset and (7) controls included and not included in the dataset.

For discrete categorical variables, Chi square or Fisher's exact tests were used. For quantitative variables, when they showed a normal distribution (total PANSS score), Student's *t* test was used; when they showed a non-normal distribution (age, duration of illness and CGAS score), non-parametric tests (Mann–Whitney *U* tests) were used. Differences of  $p < 0.05$  were considered significant. All statistical tests were two-tailed and were performed using SPSS for Windows, Version 18.0 [32].”

### Development of the SVM classifiers

Several SVM models were developed. Within each model, the output variable was membership in each stable K-SADS-based diagnostic group (SSD vs. non-SSD, SSD vs. HC, and non-SSD vs. HC). For the purposes of this study, we focused on the SSD vs. non-SSD model, as SSD and non-SSD are clinically similar and unstable diagnostic groups during the FEP. Thus, it makes sense to use this model to investigate the combination of variables at an FEP that would help us to predict a stable diagnostic outcome of SSD. As input variables, we entered a single vector with all the clinical, neuropsychological and biological variables at baseline concatenated (the ‘All variable’ classifier), or a single vector with the different set of variables concatenated (the ‘clinical’, ‘neuropsychological’, ‘neuroimaging’ and ‘biochemical’ classifiers).

First, during a training phase, only the training set was used to compute each SVM classifier. During this phase, the SVM learns from the training set and finds a discriminating pattern from each diagnostic class (e.g. SSD vs. non-SSD). To do this, the SVM draws an optimal separating hyperplane that maximizes the separation between two groups. Such a hyperplane is based on the subset of subjects lying closest to it and hence the most difficult to

classify, namely, the support vectors [26, 27]. To control possible overfitting (i.e. the inability to generalize what has been learned to novel data), several strategies were used. First, only linear kernels were used [25, 33]. Second, we performed a validation of the different classifiers using two techniques in the training set: leave one out cross-validation (LOOCV) and jackknifing. Both techniques estimate the ability of the classifier to reliably classify new cases. LOOCV is an  $N$ -fold cross-validation method that estimates the performance of a classifier by training the algorithm with  $N - 1$  subjects and classifying the  $N$ th subject repeatedly for all  $N$  subjects. LOOCV is known to provide almost unbiased estimates with moderate variance. Jackknifing repeats  $N^{\text{jackknife}}$  times the process of randomly splitting  $N$  subjects into a test set with  $N/k$  subjects and a testing set with  $(k - 1)N/k$  subjects. The estimation of the classifier performance is the average of each repetition's result. This enables a reduction of the variance of the estimation. In this study  $N^{\text{jackknife}} = 1,000$  and  $k = 5$ . The performance of each classifier was estimated in terms of accuracy (proportion of correct classifications), sensitivity (proportion of true positives correctly identified) and specificity (proportion of true negatives correctly identified).

After the training phase, the SVM classifiers were validated with the previously unseen test dataset (formed by the remaining 20 % of the original sample). The performance estimates (accuracy, sensitivity and specificity) were again calculated.

Finally to rank all the variables by their predictive weight for a particular diagnostic outcome and to improve the generalization ability of the classifier, a feature selection phase was performed after the validation phase. The feature selection phase enables the removal of irrelevant and redundant (e.g. highly correlated) features in the dataset. By selecting those features which discriminate best between groups, it increases the performance of the predictive classifier. Feature selection is particularly helpful in small sample classification problems, where the number of available training samples is very small compared to the number of features, as was the case in our study. Specifically enhanced recursive feature elimination (EnRFE), an RFE-derived technique, was applied to each of our previously validated classifiers. The basic principle of RFE is to rank and prune the available variables in the classifier using the weight of each variable in constructing the SVM decision hyperplane as a ranking criterion [34]. To do so, standard RFE removes a

feature if it is weak or redundant at a particular step and retains independent and relevant features. EnRFE is a RFE-derived technique which redefines the criterion of removing features at each state of the training so that redundant or weak features that improve performance estimates when combined between them or with other relevant features can be retained [35]. In an iterative scheme, the SVM-EnRFE classifier is trained until a core set of variables with the highest discriminative power remains. Performance estimates can be again calculated for these SVM-EnRFE classifiers through validation procedures [34]. In our study EnRFE was performed for all the previously validated SVM classifiers, with the exception of the ‘Biochemical variable’ classifier, where an EnRFE technique could not be performed since there was only one input variable (total antioxidant status—TAS, see Table 1). All SVM-EnRFE classifiers were validated using LOOCV, jackknifing and the test dataset and performance estimates were calculated for each of them. Also to assess the statistical significance of the different estimates, permutation testing was performed for each of the SVM and SVM-EnRFE classifiers validated with LOOCV and jackknifing [36]. Although the development of a single SVM classifier does not require any correction for multiple comparisons, given that multiple SVMs were generated in our study ( $n = 45$ ), critical  $p$  values were corrected for multiple comparisons using false discovery rate (FDR) type II ( $q = 0.5$ ) [37]. All analyses were performed using MATLAB Version 7.13.0.564 (R2011b) and the Statistic Toolbox Version 7.6 (R2011b) [38].

## Results

### Sample description at baseline

Patients included in the dataset ( $n = 81$ ) and patients excluded from the dataset ( $n = 29$ ) did not differ significantly in age, sex, ethnicity, total PANSS score, CGI-S score, or CGAS score at baseline. The same was true for the healthy controls included in the dataset ( $n = 42$ ) and excluded from the dataset ( $n = 56$ ) (data available upon request).

Table 2 shows the demographic and clinical characteristics at baseline of the 81 (49 SSD, 32 non-SSD) patients and 42 healthy controls who were included in the dataset and who were assigned either to the training set or to the test set before the development of the SVM classifiers. SSD and non-SSD, SSD and controls, and non-SSD and controls included in the original, training or test datasets did not differ significantly in demographic or clinical

characteristics (except for a higher CGAS score at baseline in the control groups compared with both patient groups). Neither did patients and controls included in the training set nor patients and controls included in the test set.

**Table 2**

Baseline demographic and clinical characteristics of SSD patients, non-SSD patients

	Whole dataset ( <i>N</i> = 123)			<i>p</i> values		
	SSD ( <i>N</i> = 49)	Non-SSD ( <i>N</i> = 32)	HC ( <i>N</i> = 42)	SSD vs. non- SSD	SSD vs. HC	Non- SSD vs. HC
Age (years), mean (SD) [range]	15.2 (1.9) [9–17]	15.6 (1.8) [11–17]	15.2 (1.5) [12–17]	0.36	0.46	0.09
Male sex, <i>N</i> (%)	38 (77.6)	19 (59.4)	29 (69.0)	0.08	0.36	0.39
Ethnicity (Caucasian), <i>N</i> (%)	43 (87.8)	29 (90.6)	40 (95.2)	0.99	0.28	0.65
PANSS total score, mean (SD)	89.4 (17.8)	90.4 (24.1)	–	0.83	–	–
CGAS score, mean (SD)	33.5 (14.7)	34.7 (14.9)	91.9 (4.5)	0.56	<b>&lt;0.001</b>	<b>&lt;0.001</b>
Duration of illness (days), mean (SD), [range]	70.9 (52.6) [2–180]	54.7 (47.0) [5– 180]	–	0.20	–	–
Antipsychotic treatment, <i>N</i> (%)	47 (95.9)	32 (100)	–	0.52	–	–
Subjects per site, <i>N</i> Site 1/2/3/4/5	25/8/0/3/13	16/4/5/2/5	27/1/3/3/8	0.06	0.06	0.32

In all cells, % refers to percentages (within columns) of participants. For qualitative variables that showed a normal distribution (total PANSS score), Student's *t* test was used; when not, Fisher's exact test was used. Statistically significant *p* values in bold

CGAS children's global assessment scale, PANSS positive and negative symptom



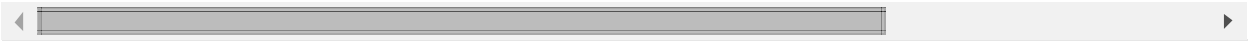
Table 3 shows the performance estimates of the SVM-EnRFE classifiers (SSD vs. non-SSD, SSD vs. HC and non-SSD vs. HC classifiers), considering all variables together or the different sets of variables, and validated with LOOCV, with jackknifing or with the test set. Among the SSD vs. non-SSD models, the highest performance estimates were obtained with the 'All variable' classifier validated with the test set (Table 3).

**Table 3**

Performance estimates of the SVM-En RFE classifiers after validation with LOOC test dataset

Baseline variables <sup>a</sup>	SSD vs. non-SSD ( <i>N</i> = 49 vs. 32)			SSD vs. HC ( <i>N</i> = 49 vs. 42)		
	acc	sens	spec	acc	sens	spec
All variables ( <i>n</i> = 1,050)						
LOOCV	0.78**	0.87**	0.65**	0.99***	0.99***	0.99***
Jackknifing	0.79***	0.87***	0.68***	0.99***	0.99***	0.99***
Test set	0.81	0.90	0.67	0.99	0.99	0.99
Clinical variables ( <i>n</i> = 747)						
LOOCV	0.91***	0.87***	0.96***	0.99***	0.99***	0.99***
Jackknifing	0.86**	0.88*	0.83**	0.99**	0.99**	0.99**
Test set	0.75	0.90	0.50	0.99	0.99	0.99
Neuropsychological variables ( <i>n</i> = 81)						
LOOCV	0.63 <sup>ns</sup>	0.74 <sup>ns</sup>	0.46 <sup>ns</sup>	0.67***	0.62***	0.74***
Jackknifing	0.63 <sup>ns</sup>	0.66 <sup>ns</sup>	0.57 <sup>ns</sup>	0.87**	0.85**	0.88**
Test set	0.56	0.70	0.33	0.83	0.80	0.88
Neuroimaging variables ( <i>n</i> = 221)						
LOOCV	0.60 <sup>ns</sup>	0.67**	0.50 <sup>ns</sup>	0.88**	0.87 <sup>ns</sup>	0.88***
Jackknifing	0.60 <sup>ns</sup>	0.67*	0.50 <sup>ns</sup>	0.67**	0.65*	0.69***
Test set	0.50	0.40	0.67	0.61	0.40	0.88
Biochemical variables ( <i>n</i> = 1) <sup>a</sup>						
LOOCV	0.45 <sup>ns</sup>	0.54 <sup>ns</sup>	0.31 <sup>ns</sup>	0.63*	0.74*	0.5 <sup>ns</sup>
Jackknifing	0.46 <sup>ns</sup>	0.56 <sup>ns</sup>	0.31 <sup>ns</sup>	0.63*	0.74**	0.5 <sup>ns</sup>

Test set	0.63	0.60	0.67	0.78	0.80	0.75	
SVM Input: baseline variables. SVM Output: diagnostic outcome (SSD vs. non-SSD vs. HC). <i>acc</i> accuracy, <i>EnRFE</i> enhanced-recursive feature elimination, <i>HC</i> healthy out cross-validation, <i>non-SSD</i> non-schizophrenia spectrum disorders, <i>sens</i> sensitivity, <i>specificity</i> specificity, <i>SVM</i> support vector machine							
<sup>a</sup> EnRFE was not performed for the ‘Biochemical variable’ classifier since the number of variables $n = 1$							
* $p < 0.05$ , ** $p < 0.01$ , *** $p < 0.001$ , <i>ns</i> non-significant. For all the LOOCV-classifiers, critical $p$ values are corrected for multiple comparisons with false discovery rate ( $q = 0.5$ )							



For the ‘All-variable-SSD vs. non SSD’ classifier, after EnRFE and validation with the test set, 243 variables (213 clinical and 30 neurocognitive variables) were selected because of their highest discriminative value for a diagnostic outcome of SSD. Table 4 shows the main first-ranked variables within this SVM-EnRFE model. Biochemical and neuroimaging variables were all ruled out by EnRFE, as they did not provide any additional predictive value within the best SSD model. The combination of 243 clinical and cognitive variables at baseline made it possible to classify patients into those who will eventually be diagnosed with SSD with an accuracy of 0.81, a sensitivity of 0.90, and a specificity of 0.67.

**Table 4**

Main baseline variables with the highest predictive weight for a diagnostic outcome of SSD

Clinical variables ( $n = 213$ )
>Severity of negative or disorganized symptoms preceding and/or during <i>FEP</i> : flattened affect, social withdrawal, stereotyped thinking and conceptual disorganization (medical records and items P2, N1, N4, N7 of the PANSS)
>Severity of hallucinations (item P3 of the PANSS)
<Insight (item G12 of the PANSS)
<Severity of affective symptoms (YMRS and HDRS total scores)
>Prescription of risperidone and aripiprazole
>Antipsychotic dose
>% of ‘true’ responses within particular FES items (e.g. item 20, 22)

<Severity of somatic complaints (items G1 of the PANSS and 13 of the HDRS) or hypochondria (item 15 of the HDRS)
>Prevalence of obstetric complications (Lewis–Murray scale of obstetric data)
Neuropsychological variables ( $n = 30$ )
<Estimated IQ
<Attention score
<Global cognition score
>Motor coordination impairment (NES motor coordination score)
Neuroimaging variables ( $n = 0$ )
None
Biochemical variables ( $n = 0$ )
None
Main first-ranked variables within the SVM-EnRFE ‘SSD vs. non-SSD’ classifier after validation with the test dataset
<i>EnRFE</i> enhanced recursive feature elimination, <i>FEP</i> first episode of psychosis, <i>FES</i> family environment scale, <i>HDRS</i> Hamilton depression rating scale, <i>IQ</i> intelligence quotient, <i>NES</i> neurological examination scale, <i>PANSS</i> positive and negative symptom scale ( <i>P</i> positive symptom subscale, <i>N</i> negative symptom subscale, <i>G</i> general symptom Subscale), <i>SSD</i> schizophrenia spectrum disorder, <i>SVM</i> support vector machine, <i>YMRS</i> young mania rating scale

## Discussion

In the present study, we found that SVMs could be helpful multivariate statistical tools in the search for predictors of diagnostic outcomes in patients with early-onset FEP. By grouping together a large amount of clinical, neuropsychological and biological data at baseline (during the FEP), we designed an SVM model that selected a combination of clinical variables (higher severity of negative symptoms and hallucinations or poorer insight) and neuropsychological variables (higher impairment in attention, global cognition, and motor coordination) as those with the highest predictive value for a diagnostic outcome of SSD. This combination of variables classified child and adolescent FEP patients who would eventually be diagnosed with SSD with an accuracy of 0.81. Total antioxidant status or neuroimaging variables such as frontal gray volume were not selected for the best SVM model as they did not provide additional predictive value. Therefore, the

assessment of particular clinical and neuropsychological variables can be helpful for guiding differential diagnosis during an early-onset FEP.

To our knowledge, this is the first study to use an SVM model as a multivariate statistical tool to identify particular variables from different sources that might predict a diagnostic outcome of SSD in the early stages of an early-onset FEP. Despite the increasing application of SVMs in medicine as tools for predicting diagnosis at early stages of disease [39–41], their application in psychosis has been limited (Orri, Petterson-Yeo et al. [42] for a review). SVMs have been used more as tools to investigate potential biomarkers for a specific diagnosis in cross-sectional studies that only included chronic, well-characterized patients. In addition, they have usually used a single source of data (mainly neuroimaging data) [19, 43–45]. SVMs have also been used to search for predictors of disease course in patients with FEP [46] and of transition to psychosis in at-risk individuals [47–49]. However, no studies had previously used SVMs to search for predictors of diagnostic outcome in early-onset FEP.

In our study, the selected predictive variables for a diagnostic outcome of SSD are to some extent consistent with those described in previous longitudinal studies in FEP patients using traditional multivariate approaches. With regard to clinical variables, history of obstetric complications [12], the presence of negative symptoms [9, 13] and poor insight [50] have been described as good predictors of diagnosis outcome or diagnostic shift to schizophrenia both in adult and adolescent samples. Furthermore positive symptoms such as hallucinations are usually more frequent and severe in SSD patients than in patients with other psychotic disorders [51]. Lower severity of depressive symptoms also predicted a diagnostic outcome of SSD in our sample, as it predicted change of diagnosis to SSD in a previous study by our group which used a clinically comparable sample of FEP patients from the CAFEPS study [52]. Although other previous longitudinal studies in FEP patients using traditional multivariate approaches have reported an association between lifetime substance disorder and change of diagnosis to SSD [9], substance use or presence of substance disorder were not selected in the best SSD vs. non-SSD predictive model, which was also congruent with a previous CAFEPS-related study [8]. Conversely with regard to family communication and environment, our study found that specific items from the family environment scale (FES) (e.g. higher percentage of ‘true’ answers for item 20: having few rules at home or item 22: struggling with venting at

home) were associated with a diagnostic outcome of SSD, whereas a previous study by our group did not find an association between family communication skills (assessed using the Parent-Adolescent Communication Inventory—PACI) and 1-year diagnosis of schizophrenia or bipolar disorder [53]. Furthermore, we found that being prescribed risperidone or aripiprazole at baseline was a good predictor of SSD diagnostic outcome, in combination with a higher severity of negative symptoms and the rest of the selected clinical and cognitive variables, whereas another CAFEPS-derived study found that clinicians seem to prescribe olanzapine and quetiapine more than risperidone when negative symptoms are prominent [54]. However, the above-mentioned studies (including previous CAFEPS-related studies) used traditional multivariate statistical approaches (e.g. logistic regression) and focused only on a specific source of measurements (i.e. clinical) and a limited number of variables (e.g. total scores or subscores of a specific scale), whereas our study relied on the simultaneous assessment of clinical, cognitive, biochemical and neuroimaging measurements at a single-patient level instead of at a group level at a very early stage of the illness. Our findings reinforce the idea that clinical prediction in psychiatry requires a combination of multiple measurements from different sources, especially in complex disorders [24].

A generalized deficit in the cognitive domain and the presence of neurological soft signs have been described in schizophrenia and other psychotic disorders such as affective psychoses [55–59], thus potentially explaining the low performance estimates of the EnRFE ‘SSD vs. non SSD’ models that included only neuropsychological variables (and the non-significance for those validated with LOOCV and jackknifing). However, within the model that used all baseline variables together, the combination of these and other neuropsychological variables with particular clinical variables easily predicted a diagnostic outcome of SSD. Again, this finding supports the need for combining multiple inputs from clinical and neuropsychological sources when predicting a diagnosis of psychosis. It is also consistent with findings from studies reporting that cognitive impairment is more severe, appears earlier, and tends to be more independent of clinical symptoms in schizophrenia than in other psychotic disorders, especially at onset [15, 60]. Contrary to our expectations [15, 16], executive function impairment was not selected among the variables with the highest predictive weight for a diagnostic outcome of SSD.

Performance estimates of the EnRFE ‘SSD vs. non-SSD’ model which included only neuroimaging variables were relatively low (and non-significant for those validated with LOOCV and jackknifing). Furthermore, within the EnRFE classifier that considered all the baseline variables together, neuroimaging variables were not selected since they did not provide any additional predictive value for the diagnostic outcome of SSD. These results were not congruent with those from previous studies including individual neuroimaging data for diagnostic prediction of SSD and other psychotic disorders [19, 45]. GM deficits in prefrontal cortex, insula, amygdala or hippocampus have been described as accurate predictors of an SSD diagnosis both in first episode and chronic patients [18–20], and higher estimates (e.g. accuracies >0.75) for neuroimaging variables alone have been reported. However, since these studies are cross-sectional and based on samples of adult patients, their results are not comparable with ours. Moreover, studies in child and adolescent psychiatric disorders using neuroimaging data and SVMs for the prediction of diagnostic outcomes usually report lower performance estimates than studies in adult samples [25]. In addition, since only volumetric measures for various regions of interest (ROIs) were included in our model, our findings are debatable, considering that the structural abnormalities that have been described in brains of patients with SSD and other psychotic disorders are subtle and actually very distributed (Yu, Cheung et al. [18] for a meta-analysis). Lastly this multicenter study had several limitations in terms of acquisition and processing procedures (see Online Resource 1 for a detailed description of the neuroimaging procedure). All this may have precluded finding neuroimaging predictors of diagnostic outcome [25].

Furthermore within the classifier that used all the baseline variables together, total antioxidant status (TAS) (the only biochemical variable) did not provide additional predictive value for a diagnostic outcome of SSD. Also performance estimates of the ‘SSD vs. non-SSD’ models which included only this biochemical variable were relatively low (and non-significant for those validated with LOOCV and jackknifing). These results are in line with those from previous studies in FEP patients where biochemical markers such as decreased TAS seem to be more associated with severity and chronicity of psychosis than with a specific diagnosis [21]. This may also be due to the fact that the other sets of variables outnumbered the biochemical set. Finally we were not able to study the predictive value of other biochemical variables such as cellular enzyme activities, lipid peroxidation or glutathione levels,

since they were excluded following our missing data strategy.

The findings from this study should be interpreted in the context of other important limitations. First, most of the patients included had acute first episodes of psychosis while hospitalized, thus precluding generalization of the findings. Second, during construction of the dataset, several subjects were excluded (e.g. those who had not undergone neuroimaging); consequently, the results could be biased, since these patients could be the most acutely or severely ill and least cooperative patients. That said, patients included in the analysis did not differ significantly from those excluded in terms of demographic data, severity of symptoms or functional impairment. Third, several variables were discarded and zero and mean imputation were performed for treating missing data. This may have affected the performance of the classifier and the generalizability of the results. Fourth, only 20 % of the sample that came from the same dataset as the training set was used for the testing phase. Fifth, there was a small imbalance in the sample size of the SSD ( $N = 49$ ), non-SSD ( $N = 32$ ) and HC ( $N = 42$ ) group. This may have artificially increased the accuracy of the SSD vs. non-SSD classifier, since it was biased toward the sensitivity estimate (Table 3). Sixth, the use of a larger sample size and a higher number or a different set of neuroimaging and biochemical variables could have increased classification accuracy and could have revealed biomarkers with higher predictive value for a diagnostic outcome of SSD. Seventh, a longer follow-up period could have enabled us to rely on more accurate gold-standard diagnostic outputs for developing the classifier, since there was a possibility that the diagnoses changed again after the 2-year observation period [61, 62]. The stability of schizophrenia is very high even in follow-up studies as long as 11 years (80 %) [61] or even 42 years (91 %) [62], but it is possible that the continuity with adult schizophrenia is not total. There may be cases where this change in diagnosis is not related to low accuracy of previous diagnosis but to “true changes”. An approach of considering schizophrenia as one clinical manifestation of long-life developmental abnormalities, with other possible clinical manifestations at other times, may also be a valid one. Last, since we included all available baseline clinical, neuropsychological, magnetic resonance imaging brain volumetric and biochemical variables in a multivariate high-dimensional model, one could argue that the study was not hypothesis-driven but exploratory. Although they might lack statistical rigor, exploratory analyses make it possible to find patterns in the data. These patterns can then be statistically tested (1) by permutation analyses, which

provide a significance value for each of the classification and (2) by dividing the original dataset into the training dataset and the previously unseen test dataset, as we did in this study.

Finally, since the clinical variables that conformed to the DSM diagnosis were included in the classifier, and since they outnumbered the other types of variables, the performance of the classifier might be expected to be excellent, with clinical variables being obviously the most predictive. Moreover, since the clinician is able to diagnose with 80 % accuracy at baseline (as around 20 % of diagnoses is unstable at this stage) and the SVM model did this with the same accuracy (81 %), one might wonder what SVM adds to assist differential diagnosis. The answer may be that it can be helpful as a statistical multivariate predictive tool in the search for sensitive and specific predictors of diagnosis rather than as a classification tool, which was the main aim of this study. Description of these predictors could then assist clinicians in terms of differential diagnosis at the FEP.

The main strengths of our study include the recruitment of child and adolescent patients at first onset, which avoids the potential confounding effect of medication and other factors related to disease progression; the relatively large sample size, the use of a ‘previously unseen’ test dataset, the wealth of data and the number of different predictors. Moreover compared with other multivariate analysis methods, SVMs offer several advantages, since they enable the assessment of high-dimensional datasets including linearly and non-linearly highly correlated variables, and the selection of a combination of variables (even if redundant or with a weak predictive weight when separately considered) that may provide the best class separation between two groups, with no need for multiple comparison correction, with minimisation of the problem of circularity and with reduction of the generalization error, especially after using EnRFE. This in turn may help devise more parsimonious clinical and neuropsychological batteries to assess patients at intake, when patients are clinically similar and their diagnoses are unstable, and thus help clinical decision making at this stage.

In summary, using SVMs, we found that a combination of clinical variables (severity of negative symptoms, hallucinations or poor insight) and neuropsychological variables (attention, global cognitive or motor coordination deficits) at the emergence of an early-onset FEP had the highest predictive value for a diagnostic outcome of SSD, as opposed to



neuroimaging and biochemical variables which did not provide additional predictive value. Hence, clinical judgment based on comprehensive clinical and cognitive assessment still seems to be the most valuable tool for guiding differential diagnosis in patients with early-onset FEP. Furthermore, current psychosis classification systems should continue to be based on traditional nosology until strong neurobiological evidence can support change through sensitive and specific biomarkers [63]. To find this evidence, SVMs could prove to be promising tools.

AQ3

## Acknowledgments

The authors would like to thank all participants and their families. The study was supported by CIBERSAM, Instituto de Salud Carlos III, the Spanish Ministry of Economy and Competitiveness, Red Temática de Investigación Cooperativa Sanitaria (RD06/0011, Red de Enfermedades Mentales y Trastornos Afectivos y Psicóticos Network); the Spanish Ministry of Health and Social Policy (Grants PI02/1248, PI05/0678, and G03/032); the CDTI under the CENIT Program (AMIT Project); Madrid Regional Government (S2010/BMD-2422 AGES) and European Union Structural Funds; Fundación Alicia Koplowitz; Fundación Mutua Madrileña; Caja Navarra; and the Network of European Funding for Neuroscience Research (ERANET-NEURON). Laura Pina-Camacho holds a grant from the Alicia Koplowitz Foundation, Spain.

### *Conflict of interest*

The authors report no conflict of interest in relation with this study.

### *Ethical standards*

The study was approved by the local institutional review boards of all the participating centers and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. All participants and their parents or legal guardians signed an informed consent form prior to their inclusion in the study.

## Electronic supplementary material

Below is the link to the electronic supplementary material.

Supplementary material 1 (DOC 41 kb)

## References

1. Catala-Lopez F, Genova-Maleras R, Alvarez-Martin E, Fernandez de Larrea-Baz N, Morant-Ginestar C (2013) Burden of disease in adolescents and young people in Spain. *Rev Psiquiatr Salud Ment* 6:80–85
2. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, Sawyer SM, Mathers CD (2011) Global burden of disease in young people aged 10–24 years: a systematic analysis. *Lancet* 377:2093–2102
3. Craddock N, Owen MJ (2010) The Kraepelinian dichotomy - going, going... but still not gone. *Br J Psychiatry* 196:92–95
4. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders. American Psychiatric Publishing, Arlington
5. American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders. Washington
6. World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization, Geneva
7. Bromet EJ, Naz B, Fochtmann LJ, Carlson GA, Tanenberg-Karant M (2005) Long-term diagnostic stability and outcome in recent first-episode cohort studies of schizophrenia. *Schizophr Bull* 31:639–649
8. Castro-Fornieles J, Baeza I, de la Serna E, Gonzalez-Pinto A, Parellada M, Graell M, Moreno D, Otero S, Arango C (2011) Two-year diagnostic stability in early-onset first-episode psychosis. *J Child Psychol Psychiatry* 52:1089–1098
9. Schwartz JE, Fennig S, Tanenberg-Karant M, Carlson G, Craig T, Galambos N, Lavelle J, Bromet EJ (2000) Congruence of diagnoses 2 years after a first-admission diagnosis of psychosis. *Arch Gen Psychiatry* 57:593–600

10. Menezes NM, Milovan E (2000) First-episode psychosis: a comparative review of diagnostic evolution and predictive variables in adolescents versus adults. *Can J Psychiatry* 45:710–716
11. Arango C, Fraguas D, Parellada M (2014) Differential neurodevelopmental trajectories in patients with early-onset bipolar and schizophrenia disorders. *Schizophr Bull* 40(Suppl 2):S138–S146
12. Cannon M, Jones PB, Murray RM (2002) Obstetric complications and schizophrenia: historical and meta-analytic review. *Am J Psychiatry* 159:1080–1092
13. McClellan J, McCurry C, Speltz ML, Jones K (2002) Symptom factors in early-onset psychotic disorders. *J Am Acad Child Adolesc Psychiatry* 41:791–798
14. Schimmelmann BG, Conus P, Edwards J, McGorry PD, Lambert M (2005) Diagnostic stability 18 months after treatment initiation for first-episode psychosis. *J Clin Psychiatry* 66:1239–1246
15. Pena J, Ojeda N, Segarra R, Eguiluz JI, Garcia J, Gutierrez M (2011) Executive functioning correctly classified diagnoses in patients with first-episode psychosis: evidence from a 2-year longitudinal study. *Schizophr Res* 126:77–80
16. Correll CU, Smith CW, Auther AM, McLaughlin D, Shah M, Foley C, Olsen R, Lencz T, Kane JM, Cornblatt BA (2008) Predictors of remission, schizophrenia, and bipolar disorder in adolescents with brief psychotic disorder or psychotic disorder not otherwise specified considered at very high risk for schizophrenia. *J Child Adolesc Psychopharmacol* 18:475–490
17. Smieskova R, Fusar-Poli P, Allen P, Bendfeldt K, Stieglitz RD, Drewe J, Radue EW, McGuire PK, Riecher-Rossler A, Borgwardt SJ (2010) Neuroimaging predictors of transition to psychosis—a systematic review and meta-analysis. *Neurosci Biobehav Rev* 34:1207–1222
18. Yu K, Cheung C, Leung M, Li Q, Chua S, McAlonan G (2010) Are Bipolar Disorder and Schizophrenia Neuroanatomically Distinct? *An*

19. Bansal R, Staib LH, Laine AF, Hao X, Xu D, Liu J, Weissman M, Peterson BS (2012) Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. *PLoS One* 7:e50698
20. Schnack HG, Nieuwenhuis M, van Haren NE, Abramovic L, Scheewe TW, Brouwer RM, Hulshoff Pol HE, Kahn RS (2014) Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage* 84:299–306
21. Mico JA, Rojas-Corrales MO, Gibert-Rahola J, Parellada M, Moreno D, Fraguas D, Graell M, Gil J, Irazusta J, Castro-Fornieles J, Soutullo C, Arango C, Otero S, Navarro A, Baeza I, Martinez-Cengotitabengoa M, Gonzalez-Pinto A (2011) Reduced antioxidant defense in early onset first-episode psychosis: a case-control study. *BMC Psychiatry* 11:26
22. Kapur S, Phillips AG, Insel TR (2012) Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry* 17:1174–1179
23. Van Os J, Allardyce J (2009) The clinical epidemiology of schizophrenia. In: Kaplan B, Saddock J, Saddock V, P R (eds) *Kaplan and Saddock's comprehensive textbook of psychiatry*. Lippincott Williams Wilkins, London
24. Sperling R, Johnson K (2013) Biomarkers of Alzheimer disease: current and future applications to diagnostic criteria. *Continuum (Minneapolis)* 19:325–338
25. Johnston BA, Mwangi B, Matthews K, Coghill D, Steele JD (2013) Predictive classification of individual magnetic resonance imaging scans from children and adolescents. *Eur Child Adolesc Psychiatry* 22:733–744
26. Vapnik VN (1998) *Statistical learning theory*. Wiley, New York
27. Boser BE, Guyon I, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Haussler D (ed) *5th Annual ACM*

Workshop on COLT. ACM Press, Pittsburgh, pp 144–152

28. Steeking R, Schebesch KB (2003) Support Vector Machines for Credit Scoring: Comparing to and Combining With Some Traditional Classification Methods. In: Schader M, Gaul W, Vichi M (eds) *Between Data Science and Applied Data Analysis*. Springer, Berlin, pp 604–612
29. Castro-Fornieles J, Parellada M, Gonzalez-Pinto A, Moreno D, Graell M, Baeza I, Otero S, Soutullo CA, Crespo-Facorro B, Ruiz-Sancho A, Desco M, Rojas-Corrales O, Patino A, Carrasco-Marin E, Arango C (2007) The child and adolescent first-episode psychosis study (CAFEPS): design and baseline results. *Schizophr Res* 91:226–237
30. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N (1997) Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36:980–988
31. Bishop CM, N.M. N (2006) *Pattern recognition and machine learning*. Springer, New York
32. SPSS Inc (2009) PASW statistics for windows. In: SPSS Inc., Chicago
33. Abu-Mostafa YS, Magdon-Ismail M, Lin HT (2012) *Learning from Data: A short Course*. AML Book, Pasadena
34. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *Journal of Machine Learning Research* 3:1157–1182
35. Chen X, Jeong J (2007) Enhanced recursive feature elimination. In: *Machine learning and applications, 2007. ICMLA 2007. Sixth International Conference on Machine learning and applications*. IEEE, Cincinnati, pp 429–435
36. Noirhomme Q, Lesenfans D, Gomez F, Soddu A, Schrouff J, Garraux G, Luxen A, Phillips C, Laureys S (2014) Biased binomial assessment of cross-validated estimation of classification accuracies

illustrated in diagnosis predictions. *Neuroimage Clin* 4:687–694

37. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 51:289–300

38. The MathWorks Inc. (2011) MATLAB and statistics toolbox release 2011b. In, Natick, Massachusetts

39. Al-Naami B, Al-Nabulsi J, Amasha H, Torry J (2010) Utilizing wavelet transform and support vector machine for detection of the paradoxical splitting in the second heart sound. *Med Biol Eng Comput* 48:177–184

40. Kim N, Seo JB, Lee Y, Lee JG, Kim SS, Kang SH (2009) Development of an automatic classification system for differentiation of obstructive lung disease using HRCT. *J Digit Imaging* 22:136–148

41. Prilutsky D, Shneider E, Shefer A, Rogachev B, Lobel L, Last M, Marks RS (2011) Differentiation between viral and bacterial acute infections using chemiluminescent signatures of circulating phagocytes. *Anal Chem* 83:4258–4265

42. Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A (2012) Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* 36:1140–1152

43. Costafreda SG, Fu CH, Picchioni M, Touloupoulou T, McDonald C, Kravariti E, Walshe M, Prata D, Murray RM, McGuire PK (2011) Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry* 11:18

44. Pettersson-Yeo W, Benetti S, Marquand AF, Dell'acqua F, Williams SC, Allen P, Prata D, McGuire P, Mechelli A (2013) Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high-risk and first-episode psychosis at the individual level. *Psychol Med*:1–16

AQ4

45. Schnack HG, Nieuwenhuis M, van Haren NE, Abramovic L, Scheewe TW, Brouwer RM, Hulshoff Pol HE, Kahn RS (2013) Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage* 84C:299–306
46. Mourao-Miranda J, Reinders AA, Rocha-Rego V, Lappin J, Rondina J, Morgan C, Morgan KD, Fearon P, Jones PB, Doody GA, Murray RM, Kapur S, Dazzan P (2012) Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. *Psychol Med* 42:1037–1047
47. Koutsouleris N, Gaser C, Patschurek-Kliche K, Scheuerecker J, Bottlender R, Decker P, Schmitt G, Reiser M, Moller HJ, Meisenzahl EM (2012) Multivariate patterns of brain-cognition associations relating to vulnerability and clinical outcome in the at-risk mental states for psychosis. *Hum Brain Mapp* 33:2104–2124
48. Koutsouleris N, Davatzikos C, Bottlender R, Patschurek-Kliche K, Scheuerecker J, Decker P, Gaser C, Moller HJ, Meisenzahl EM (2012) Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophr Bull* 38:1200–1215
49. Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, Schmitt G, Zetzsche T, Decker P, Reiser M, Moller HJ, Gaser C (2009) Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry* 66:700–712
50. Cuesta MJ, Peralta V, Zarzuela A (2000) Reappraising insight in psychosis. Multi-scale longitudinal study. *Br J Psychiatry* 177:233–240
51. Lawrie SM, Olabi B, Hall J, McIntosh AM (2011) Do we have any solid evidence of clinical utility about the pathophysiology of schizophrenia? *World Psychiatry* 10:19–31
52. Nuevo R, Chatterji S, Verdes E, Naidoo N, Arango C, Ayuso-Mateos JL (2012) The continuum of psychotic symptoms in the general

population: a cross-national study. *Schizophr Bull* 38:475–485

53. Otero S, Moreno-Iniguez M, Paya B, Castro-Fornieles J, Gonzalez-Pinto A, Baeza I, Mayoral M, Graell M, Arango-Lopez C (2011) Twelve-month follow-up of family communication and psychopathology in children and adolescents with a first psychotic episode (CAFEPS study). *Psychiatry Res* 185:72–77

54. Castro-Fornieles J, Parellada M, Soutullo CA, Baeza I, Gonzalez-Pinto A, Graell M, Paya B, Moreno D, de la Serna E, Arango C (2008) Antipsychotic treatment in child and adolescent first-episode psychosis: a longitudinal naturalistic approach. *J Child Adolesc Psychopharmacol* 18:327–336

55. Reichenberg A, Harvey PD, Bowie CR, Mojtabai R, Rabinowitz J, Heaton RK, Bromet E (2009) Neuropsychological function and dysfunction in schizophrenia and psychotic affective disorders. *Schizophr Bull* 35:1022–1029

56. Hill SK, Reilly JL, Harris MS, Rosen C, Marvin RW, Deleon O, Sweeney JA (2009) A comparison of neuropsychological dysfunction in first-episode psychosis patients with unipolar depression, bipolar disorder, and schizophrenia. *Schizophr Res* 113:167–175

57. Stefanopoulou E, Manoharan A, Landau S, Geddes JR, Goodwin G, Frangou S (2009) Cognitive functioning in patients with affective disorders and schizophrenia: a meta-analysis. *Int Rev Psychiatry* 21:336–356

58. Bombin I, Arango C, Buchanan RW (2005) Significance and meaning of neurological signs in schizophrenia: two decades later. *Schizophr Bull* 31:962–977

59. Whitty P, Clarke M, McTigue O, Browne S, Gervin M, Kamali M, Lane A, Kinsella A, Waddington J, Larkin C, O'Callaghan E (2006) Diagnostic specificity and predictors of neurological soft signs in schizophrenia, bipolar disorder and other psychoses over the first 4 years of illness. *Schizophr Res* 86:110–117



60. Keefe RS, Fenton WS (2007) How should DSM-V criteria for schizophrenia include cognitive impairment? *Schizophr Bull* 33:912–920
61. Hollis C (2000) Adult outcomes of child- and adolescent-onset schizophrenia: diagnostic stability and predictive validity. *Am J Psychiatry* 157:1652–1659
62. Remschmidt H, Martin M, Fleischhaker C, Theisen FM, Hennighausen K, Gutenbrunner C, Schulz E (2007) Forty-two-years later: the outcome of childhood-onset schizophrenia. *J Neural Transm* 114:505–512
63. Miller G (2010) Psychiatry. Beyond DSM: seeking a brain-based classification of mental illness. *Science* 327:1437